Forecasting Depression in Bipolar Disorder using Cellphone Telemonitoring

Paul J. Moore*, Max A. Little, Patrick E. McSharry, *Senior Member, IEEE*, John R Geddes, Guy M Goodwin,

***Abstract*—Bipolar disorder is characterized by recurrent episodes of mania and depression and affects about 1% of the adult population. The condition can have a major impact on an individual's ability to function and is associated with a long term risk of suicide. In this paper we report on the use of self-rated mood data to forecast the next week's depression ratings. The data used in the study has been collected using SMS text messaging and comprises one time series of approximately weekly mood ratings for each patient. We find a wide variation between series: some exhibit a large change in mean over the monitored period and there is a variation in correlation structure. Almost half of the time series are forecast better by unconditional mean than by persistence. Two methods are employed for forecasting: exponential smoothing and Gaussian process regression. Neither approach gives an improvement over a persistence baseline. We conclude that the depression time series from patients with bipolar disorder are very heterogenous and that this constrains the accuracy of automated mood forecasting across the set of patients. However the dataset is a valuable resource and work remains to be done that might result in clinically useful information and tools.**

***Index Terms*—Time series analysis, Gaussian processes, Autoregressive processes, Psychiatry, Public healthcare, Health information management**

I. Introduction

BIPOLAR disorder is a condition affecting mood and featuring episodes of *mania* and *depression* which may be severe in intensity. Mania is a condition in which the sufferer might experience racing thoughts, impulsiveness, grandiose ideas and delusions. Under these circumstances, individuals are liable to indulge in activities which can be damaging both to themselves and to those around them. Depression is characterized by low mood, insomnia, problems with eating and weight, poor concentration, feelings of worthlessness, thoughts of death or suicide, a lack of general interest, fatigue and restlessness.

Effective mood forecasting in bipolar disorder could provide an early warning for signs of relapse and support management of the disorder by both clinician and patient. The symptoms have typically been monitored using paper diaries or by asking the patient to recall mood at an outpatient appointment. Partly as a result of these collection methods, modeling of mood in bipolar patients has been limited by a scarcity of suitable data. Suitability in this context implies a usable format, that is, numerical time series data with a frequency and quantity high enough for analysis. One theoretical study [1] describes a mathematical model but does not use clinical data, though the authors expressly wish to encourage its collection. Other work has been inspired by this model [2], [3]. There have also been attempts to distinguish mood episodes using nonlinear measures [4] and investigations into the dynamics of mood change in bipolar disorder [5], [6].

The paucity of suitable data has also constrained quantitative studies of mood disorders. Until recently the primary measures used have been the mean and standard deviation of the ratings from questionnaires [7], although other measures have been used: Pincus [8] has introduced *approximate entropy* as a measure of the regularity or predictability of a time series. It is useful for relatively small datasets and has since been applied to both mood data generally [9] and to mood in bipolar disorder [4]; in the latter case, 60 days of mood data from 49 patients was used for the analysis. Gottschalk *et al*. [5] analysed daily mood records from 7 rapid cycling patients with bipolar disorder and 28 normal controls. The participants in this study kept mood records on a daily basis over a period of 1 to 2.5 years. The mood charts were evaluated for periodicity and correlation dimension but the authors did not attempt to develop a complete mathematical model of mood in bipolar disorder.

At the University of Oxford, UK, GSM text messaging has been employed for telemonitoring of mood. Each week, patients complete a questionnaire and return the results as a digit sequence in an SMS text message. The resulting time series of mood ratings are visualized as color-coded graphs for use at an outpatient appointment. This information is used both by clinicians to select appropriate interventions and by the patients themselves for management of their condition. The system has generated a large database of mood time series which has the potential for research and clinical innovations [10], [11].

In this study, we apply time series forecasting methods to the data, using in-sample forecasting to examine the

*Corresponding author

P. J. Moore is with the Oxford Centre for Industrial and Applied Mathematics (OCIAM), Mathematical Institute, University of Oxford, Oxford, UK 24-29 St Giles, Oxford, OX1 3LB, UK. (phone 0044 1865283884).

M. A. Little is a Wellcome Trust/MIT Fellow with the MIT Media Lab (hd.media.mit.edu) and a senior research associate with Oxford University Systems Group.

P. E. McSharry is with the Smith School of Enterprise and the Environment. Hayes House, 75 George Street, Oxford OX1 2BQ, UK.

G. M. Goodwin is W. A. Handley Professor of Psychiatry with the Department of Psychiatry, University of Oxford, Oxford, UK.

J. R. Geddes is Professor of Epidemiological Psychiatry and Head of the Department of Psychiatry at the University of Oxford, Oxford, UK.

dynamics, and out of sample forecasting to estimate the expected prediction error. In the next section we provide a short history of time series forecasting and Gaussian process regression. In section III we describe the data set and its characteristics, including nonstationarity and smoothness. In section IV we describe the forecasting experiments and in section V we draw some conclusions from the results. The work reported here has been performed in accordance with Declaration of Helsinki of 1975, as revised in 2004.

II. Background

Stochastic time series forecasting can be traced back at least to Yule [12] who postulated that a individual time series can be regarded as the realization of a stochastic process. The autoregressive (AR) and moving average (MA) models are based on this concept. In 1970, Box and Jenkins published the influential text *Time Series Analysis: Forecasting and Control*, in which they outlined a three stage, iterative process for model identification, estimation and verification. ARIMA models and their variants have been widely applied because the behaviour of a diverse range of time series can be described using relatively few parameters. Exponential smoothing methods originated in the post-war decades with the work of Brown [13], Holt [14] and Winters [15]. They were, until the 1980s, often considered to be a collection of ad hoc techniques for univariate extrapolation but have since their inception been progressively underpinned by a statistical foundation. Muth [16] demonstrated that simple exponential smoothing (SES) gave the optimal forecast for a random walk with additive noise. Box and Jenkins [17] showed that simple exponential smoothing was equivalent to an ARIMA(0,1,1) model. Work on exponential smoothing has continued, including taxonomies of method and state space interpretations [18]. Regime switching models, such as self-exciting threshold AR (SETAR) have been introduced by Tong [19] and, in the context of the current application, were applied to bipolar depression data by Bonsall *et al.*. [11]. A review of linear and nonlinear forecasting methods are given by De Gooijer and Hyndman [20], and De Gooijer and Kumar [21].

Gaussian process regression

In most applications of time series models, the response values are assumed to be indexed at equally spaced intervals, and missing or unevenly sampled data are dealt with by a preprocessing step. Preprocessing methods include imputation for missing data and interpolation for uneven spacing, both of which rely on assumptions about the latent function. Gaussian process modeling does not require equally spaced intervals, and makes use of the time indices in learning the model. The theory for Gaussian process models was established in the 1940s and in recent decades it has become popular in regression and machine learning. O'Hagan [22] applies the model to several one-dimensional regression tasks. Williams and Rasmussen describe model optimisation [23], and Rasmussen relates Gaussian processes to other models that are commonly used in machine learning [24]. Applications in biomedical engineering have included heart rate inference [25], neuroimaging inference [26] and neonatal seizure detection [27]. The method uses a Bayesian nonparametric model and assumes a Gaussian prior distribution over the regression function. The prior is specified by a covariance function $k(\mathbf{x}, \mathbf{x}'|\,\boldsymbol{\theta})$ which defines the correlation between latent function values at inputs $\mathbf{x}$ and $\mathbf{x}'$. The *hyperparameters* $\boldsymbol{\theta}$ determine the properties, such as length scale, of the prior process and are estimated from the data by maximum likelihood along with a noise term σ_n. The expected value of the latent function is than be found from the predictive equation $\mathbb{E}[f_*] = \mathbf{k}_*^{\mathbf{T}}(K+\sigma_n^2 I)^{-1}\mathbf{y}$ where $\mathbf{k}_*$ is the vector of covariances of the test point with the training points $\mathbf{y}$, and K is the covariance matrix of the training set.

Forecasting algorithm

Table I summarizes the steps needed for forecasting by Gaussian process regression. The training set $\mathbf{y}$ with time indices $\mathbf{x}$ is first centred because the prior process is assumed to have zero mean. The algorithm then finds the optimal value for the hyperparameters $\boldsymbol{\theta}$ and the noise variance σ_n^2 by maximizing the marginal likelihood $p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$. Since this likelihood is a Gaussian distribution, its gradient with respect to $\boldsymbol{\theta}$ can be found analytically and optimized using a conjugate gradient method. The predictive equation is used to find the forecast mean and finally the original signal bias is added to give the predicted value of the mood rating.

Centre responses	$\mathbf{y} = \mathbf{y}_{in} - \mathbb{E}[\mathbf{y}_{in}]$
Find parameters $\boldsymbol{\theta}, \sigma_n^2$	Maximize $p(\mathbf{y}\|\mathbf{x}, \boldsymbol{\theta})$ wrt $\boldsymbol{\theta}, \sigma_n$
Predict response	$\mathbb{E}[f_*] = \mathbf{k}_*^{\mathbf{T}}(K + \sigma_n^2 I)^{-1}\mathbf{y}$
Add mean to estimate	$\hat{y} = \mathbb{E}[f_*] + \mathbb{E}[\mathbf{y}_{in}]$

TABLE I
PREDICTION USING GAUSSIAN PROCESS REGRESSION.

III. Data

The data used in this study is from 153 patients with bipolar disorder who were using the monitoring system between December 2006 and August 2011. Demographic data on individual patients' age is available for 119 patients and gender for 120 patients. Ages range from 20 to 75 years and for those for whom the gender was available, 47 are male and 73 female. Mood data is returned approximately each week and comprises the answers to standard self-rating questionnaires for depression and mania. The rating scale used for depression is the *Quick Inventory of Depressive Symptomatology - Self Report (QIDS-SR$_{16}$)* [28] which comprises 16 questions. This scale assesses the nine symptom domains for depression (Diagnostic and Statistical Manual of Mental Disorders, 4th edition, Text Revision) [29]. Each inventory category can contribute up to 3 points and the maximum score for each of the 9 domains is totalled, giving a total possible score of 27 on the scale. The severity of mania is quantified using the *Altman Self-Rating Mania Scale (ASRM)* [30]. This scale consists of 5 items, each of which may can contribute up to 4 points, giving a total possible score of 20 on the scale. Examples of depression and mania time series for a single patient are shown in Figure 1.

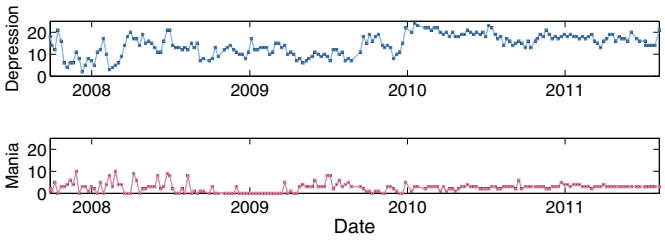


Fig. 1. Sample time series from one of the patients represented in the data set. The top plot shows the depression (QIDS) rating over a three and a half year period. The maximum possible score on the QIDS rating scale is 27. The lower plot shows the corresponding mania (ASRM) ratings, for which the maximum possible score is 20.

Data characteristics

The data set is examined by considering the time series length and the intervals between responses for each patient, and presenting these statistics for the set of patients. Figure 2 shows time series length as measured by the number of responses, plotted against the date that the patient's monitoring started. In general, patients who have enrolled

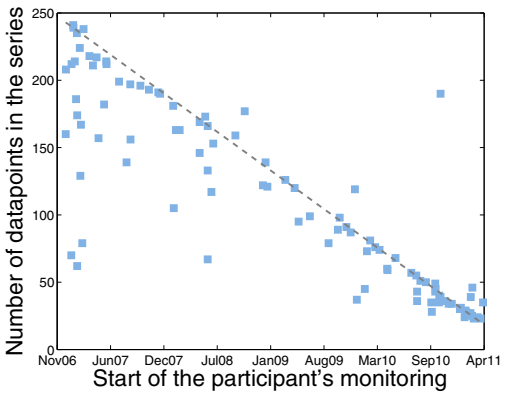
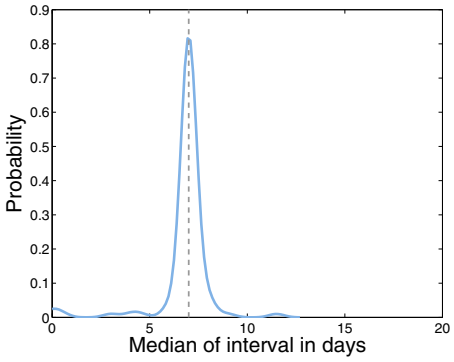


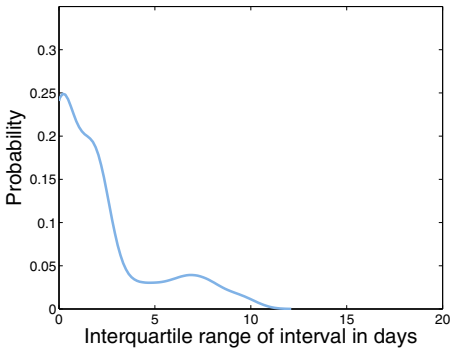
Fig. 2. A scatter plot of time series length for all patients in the data set where we define length as measured by the number of responses in the individual time series. The x-axis is the time when the patient enrolled in the monitoring system. The length of the time series is usually determined by how long the patient has been monitored: a dashed 'compliance' line shows a rate of one response per week.

more recently have shorter time series: for example those who started since the beginning of December 2010 have time series with around 30 points. Those with fewer than 23 responses are removed from the database leaving a total of 100 out of 153 patients who were in the original data set.

The statistics of the response interval for each patient are derived for the 100 patients who fulfill the minimum length criterion. Figure 3(a) shows a probability density estimate for the distribution of the median response interval per individual. Some patients return ratings more frequently than each week – often these ratings are repeated values which have been resent – while others fail to respond for one week or for a longer period. Figure 3(b) shows the distribution of the interquartile range of the intervals over the set of patients.



(a) Median interval



(b) Interquartile range of interval

Fig. 3. Response interval statistics for the set of 100 patients having 23 or more data points in their time series. Figure 3(a) is a probability density estimate for the distribution of the median time interval between ratings. The median time interval is calculated for each patient, and the estimated distribution over the patient set is shown. A vertical compliance line is drawn at a median interval of 7 days. Figure 3(b) shows the interquartile range of this interval over the set of patients. The density estimates are computed using a Gaussian kernel smoother at 100 points.

Stationarity and smoothness

We examine some properties of the time series focusing on the depression ratings because for some patients the mania scores are always zero. Some of the series exhibit a global trend, which may be the effect of medication. A quantitative test for stationarity is made by comparing the sample means for the first and second half of each patient's series along with standard errors. This test is a simplified version of a procedure described by Kantz and Schreiber [31] in which a running mean over a number of segments is used. The result, shown in Figure 4, illustrates that some patients undergo a large change in mean depression over the period of observation.

Another quality that appears to vary among patients is the roughness of the time series, a quality related to its autocorrelation function. Detrended fluctuation analysis, introduced by Peng [32], provides a measure of roughness by estimating the statistical self affinity of a time series and it allows the detection of long range correlations in the presence of nonstationarity. The algorithm finds the change in fluctuation with window size over the integrated series giving the *scaling exponent* which measures the self-affinity or smoothness of a time series. The fluctuation L is found

by first fitting a trend line within a given window length and deriving the slope m and intercept b by least squares. The fluctuation for the window size is then found as

$$F(L) = \left[\frac{1}{L} \sum_{n=1}^{L} (y(n) - mn - b^2) \right]^{\frac{1}{2}} \quad (1)$$

Self similarity is manifested as $F(L) \propto Ls$, and the scaling exponent s is found as the slope of the log-log graph of L against $F(L)$. A scaling exponent of $s = 0.5$ represents white noise, a value of 1.0 corresponds to $1/f$ noise, and a value of 1.5 corresponds to a random walk.

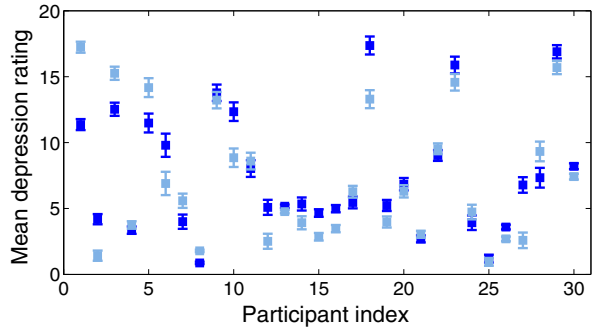


Fig. 4. Change in mean depression over the observation period. This figure shows the percentage change in mean rating of the first half (dark) of a patient's depression time series compared with the second half (light). The error bars represent the standard error assuming that the numbers are Gaussian distributed and uncorrelated. In this figure, the x-axis is a patient index: results from only the first 30 patients in the data set are shown for clarity.

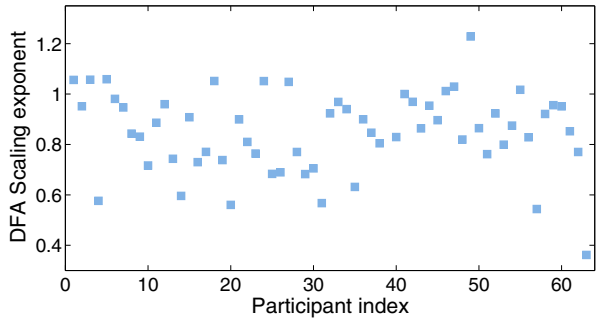


Fig. 5. Variation in smoothness over the set of time series as measured by the scaling exponent derived from detrended fluctuation analysis (DFA). Higher values of the scaling exponent correspond to smoother time series. The DFA computation is unreliable for short series, so only time series with at least 64 data points are shown.

An efficient algorithm for computing DFA is described in [33]. The performance of DFA in general has been studied under different conditions of nonstationarity and time series length [34]. In the light of these results the algorithm was applied to each of the time series in a reduced data set comprising time series having at least 64 points, and the results are shown in Figure 5.

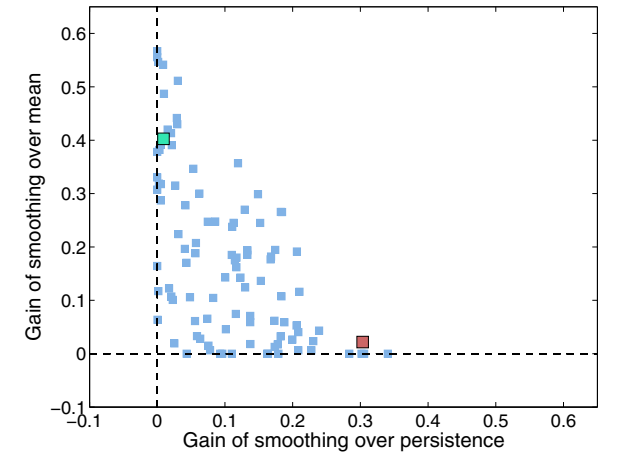
Inspection of the time plots corresponding to large s shows that those patients with $s > 1$ are highly nonstationary. For those time series with small s, the time plots show much less of a change of mean rating, and have an erratic, almost random variation. We find that few time series are periodic, many show a global trend and there is a variation from noise to random walk dynamics.

IV. Forecasting

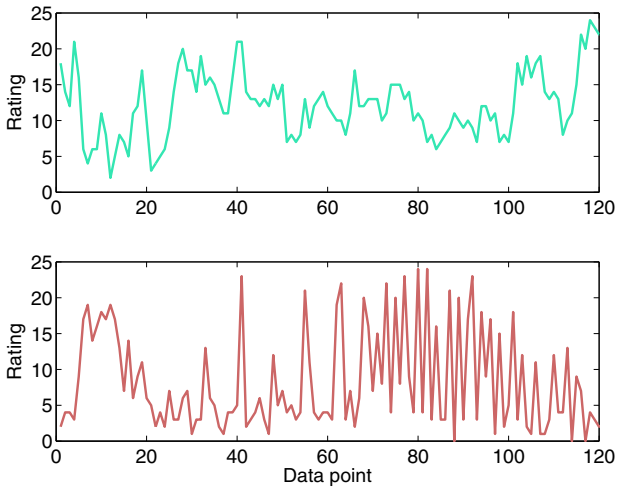
The purpose of forecasting is twofold, firstly to examine the dynamics of individual time series and secondly to estimate the expected prediction error. Time series dynamics are investigated by using in-sample forecasting and comparing the results from exponential smoothing with baseline methods. The expected prediction error is estimated using out of sample forecasting with exponential smoothing and Gaussian process regression. For both approaches, we forecast the next mood rating starting from a small margin and continuing to the end of the series. The error for the patient is then summarized by applying a loss function to the residuals, and these errors are examined for the set of patients.

For the in-sample forecasting, we compare simple exponential smoothing forecasts with baseline predictors of unconditional mean and persistence. Exponential smoothing takes a forecast $\hat{y}_t$ at time t and adjusts it to give a next step forecast of $\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t)$, where y_t is the actual value at time t and α is a constant between zero and one. In order to measure performance against both baselines simultaneously we set the seed forecast value $\hat{y}_1$ to the unconditional mean. A value of zero for the smoothing parameter then uses the unconditional mean predictor while a value of one gives a persistence forecast. The smoothing parameter is found for each patient by searching for the minimum forecast error over the parameter range of zero to one. In this way we can get some insight into the relative accuracy gains over these two baselines. Figure 6(a) shows the results displayed as a scatter plot of time series with the relative gains over the baseline methods on each axis. Gains of exponential smoothing over unconditional mean are as much as 0.57, that is, a 57% reduction in error, and 0.34 over persistence. Figure 6(b) shows examples of time plots for series where no gain is found over persistence (top panel) and none over unconditional mean (bottom panel). The lower plot varies considerably from week to week while the top plot shows less variation.

We next consider forecast results for two groups which represent these examples. One group comprises 23 patients with 'rough' time series for which unconditional mean forecasting improves by 10% or more over persistence forecasting. The second group is of 41 patients with 'smooth' time series where persistence improves by 10% or more over unconditional mean forecasts. The difference in smoothness between the sets is reflected in the DFA scaling exponents: for the smooth set the median exponent is 0.73 and for the rough set it is 0.98. Figure 7 shows the forecast error plotted against first order correlation for two groups of patients. Forecast accuracy for the groups is seen to be promoted by

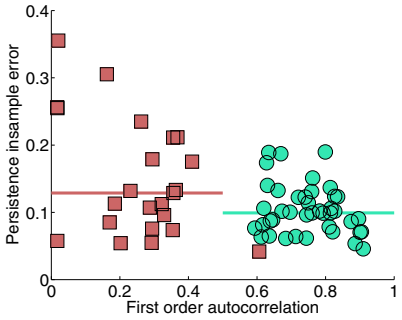


(a) Gain of simple exponential smoothing

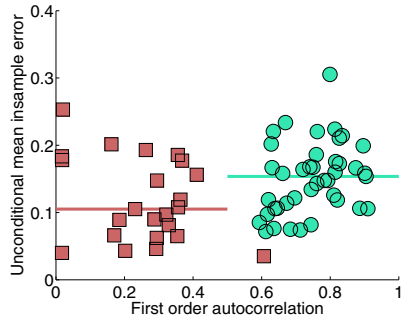


(b) Examples of time series

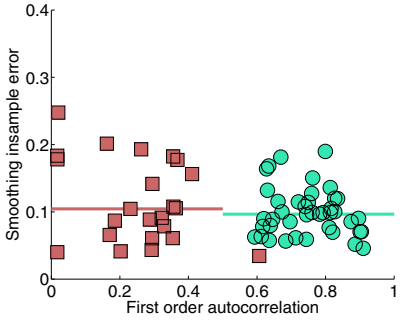
Fig. 6. The top panel shows the relative error reduction of simple exponential smoothing compared with persistence (x-axis) and unconditional mean (y-axis) where each square corresponds to a patient's time series. The gain on each axis is represented as the proportionate decrease in patient error compared with the baseline. Points in the top left region have a smoothing parameter close to 1, corresponding to a persistence forecast. Points in the bottom right region have a smoothing parameter close to 0, corresponding to a mean forecast. Simple exponential smoothing always performs at least as well as the baseline methods, but for many time series there is no improvement over the relevant baseline. Only 20 of the 100 series improve by more than 10% over both persistence and unconditional mean forecasts. The middle and lower panels are time plots corresponding to the highlighted points on the top panel and illustrate smooth and rough series respectively.



(a) Persistence



(b) Unconditional mean



(c) Exponential smoothing

Fig. 7. Forecast error against first order correlation for persistence, unconditional mean and exponential smoothing methods. Two groups of patients are shown for each method: squares correspond to rough time series and the circles to smooth time series. A horizontal line is drawn through each cluster to represent its median. The top panel shows that errors for persistence forecasting are higher and more dispersed for the rough set than for the smooth set. The situation is reversed for unconditional mean forecasting, shown in the middle panel, where the rough set shows lower errors than the smooth set. Simple exponential smoothing, shown in bottom panel, adapts to the qualities of both sets and preserves the accuracy of the respective baselines.

their relevant baseline method and penalized by the opposite method. Simple exponential smoothing preserves accuracy in both groups by adapting to the correlation structure of each time series. Most of the gain in accuracy from exponential smoothing is derived from the smoothing effect on the rough time series.

Autoregressive model orders

To investigate further the qualities of the smooth time series, we estimate an autoregressive model and use an information criterion to select the order. The optimal order $\hat{p}$ is found by minimizing the Schwarz Bayesian Criterion over a range of values of p. Out of a total of 60 time series where persistence is a better predictor better than mean, 40 are best represented by AR(1), 10 by AR(2) and the rest by orders up to 10. This result accords with the relatively poor performance of exponential smoothing compared with persistence even on series where it is a better predictor than the unconditional mean.

Out of sample forecasting

The in-sample forecasting approaches have revealed some of the dynamic properties of the time series. Out of sample forecasting gives an estimation of the expected prediction error. For this work we add Gaussian process as a forecasting method, using a rational quadratic covariance function $k(r) = \left(1 + \frac{r^2}{2\alpha l^2}\right)^{-\alpha}$ where r is the difference between

time indices, l is a length scale and α determines how the covariance changes with r. We also show the results for simple exponential smoothing and the unconditional mean and persistence baseline. The results from autoregressive models were found to be similar to those from exponential smoothing and we note that latter method has fewer parameters and so is less subject to overfitting to the training data.

Parameter values are trained on the first twenty data points of the individual time series, and retrained at intervals of ten points as the forecasts are made through the series. The results are shown in Figure 8. The median forecast patient error for the unconditional mean forecast method are slightly higher than that for the persistence, exponential smoothing and Gaussian process methods. Persistence forecasting generates the most outliers of these three methods, showing that exponential smoothing and Gaussian process methods reduce some large patient errors.

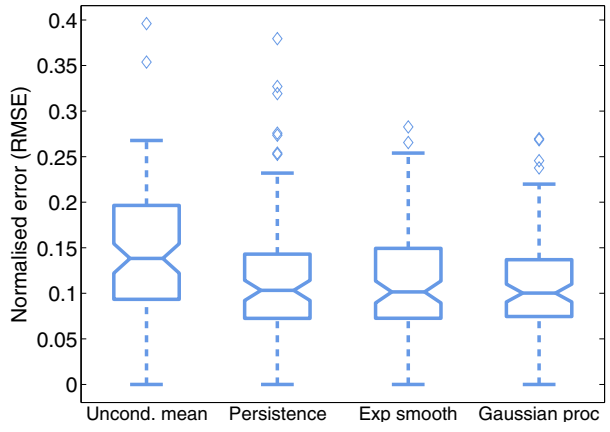


Fig. 8. Box plots for out of sample results with outliers marked by a diamond. The distributions shown are those of the errors of 100 patients, where the error for each patient is the root mean square of the next step forecasts for that patient (the number of forecasts varies from patient to patient because the time series length varies over the set). Points are drawn as outliers if they are larger than $q3 + 1.5 * (q3 - q1)$ or smaller than $q1 - 1.5(q3 - q1)$, where $q1$ and $q3$ are the 25^{th} and 75^{th} percentiles respectively.

Table II presents the numerical results for different loss functions. Results shown are the median of the distribution of errors over the patient set. Each patient's error is estimated by applying one of three loss functions to the next step forecasts: root mean square error (RMSE), mean absolute error (MAE) and median absolute error (MdAE), and errors are normalized by the maximum of the rating scale (0-27). The choice of loss function makes little difference to the relative performance of the methods, except for MdAE which gives a relatively better score to unconditional mean forecasts. In this case, it is likely that the very high residuals that characterize this baseline are not penalised as much as with the other loss functions.

Loss fn.	*Uncond.*	*Persist.*	*Exp. smooth.*	*Gauss. proc.*
RMSE	0.14	0.10	0.10	0.10
MAE	0.10	0.07	0.07	0.08
MdAE	0.07	0.06	0.06	0.07

TABLE II
OUT OF SAMPLE NEXT STEP FORECASTING RESULTS.

V. DISCUSSION

An important feature of this data set is its heterogeneity. Time series returned by the patients differ markedly in length, response interval, self-affinity and stationarity. In-sample forecasting reveals that 40 of the 100 time series are forecast better by unconditional mean than by a persistence predictor. These time series show a rapid change in rating from week to week either because of a high signal to noise ratio or undersampling: the mood is changing by day, rather than by week. The remaining 60 series that are better forecast by persistence are represented mostly by AR(1) processes, when assumed to be autoregressive models.

Simple exponential smoothing will always perform at least as well as unconditional mean or persistence forecasts for in-sample forecasting. However, only 20 of the original 153 series improve by more than 10% over both persistence and unconditional mean forecasts. Those time series that are noisy or undersampled will train a smoothing coefficient of zero and will not be improved by exponential smoothing. It is not surprising then, that exponential smoothing does not improve over the baselines for out of sample forecasting. Most of the time series are noisy or lack serial correlation, and gains from the minority of time series that benefit from smoothing will be masked by errors in other patients.

We conclude that for some patients, effective depression forecasts cannot be made over the period of a week because their time series exhibit little serial correlation. Other patients can be forecast to a degree of accuracy, but the clinical benefit is not clear because they form a minority of the whole set. An increase in accuracy might be gained by partitioning the dataset and forecasting different groups separately but it seems unlikely that the improvement would be more than marginal when averaged over the whole set. However, the mania time series remain unexplored, and although these seem less amenable than the depression series, it would be worth summarising their qualities. There is also the potential to use unsupervised methods such as clustering to identify groups of patients within the set and compare these with more detailed demographic and diagnostic information. Finally, some use might be made of the time taken for the patients to respond to prompt messages and the statistics of missing data, if allowance can be made for the SMS latency. The dataset is a valuable resource and more work remains to be done that might result in clinically useful information and tools.

ACKNOWLEDGMENT

The authors would like to acknowledge financial support from the John Fell Fund, Oxford University.

REFERENCES

[1] D. Daugherty, T. Roque-Urrea, J. Urrea-Roque, J. Troyer, S. Wirkus, and M. A. Porter, "Mathematical models of bipolar disorder," *Communications in Nonlinear Science and Numerical Simulation*, vol. 14, no. 7, pp. 2897–2908, Jul 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1007570408003705

[2] H. Mobaji and K. Karahalios, "HCI applications for aiding children with mental disorder," *Crossroads, ACM Student Magazine*, vol. 12(2), 2005.

[3] M. Anand, "Nonlinear dynamic system model of bipolar mood disorder," *Proc. IEEE, SoutheastCon, 2007.*, pp. 279 – 282, 2007.

[4] T. Glenn, P. C. Whybrow, N. Rasgon, P. Grof, M. Alda, C. Baethge, and M. Bauer, "Approximate entropy of self-reported mood prior to episodes in bipolar disorder," *Bipolar Disorders*, vol. 8, no. 5p1, pp. 424–429, Oct 2006. [Online]. Available: http://doi.wiley.com/10.1111/j.1399-5618.2006.00373.x

[5] A. Gottschalk, M. S. Bauer, and P. C. Whybrow, "Evidence of chaotic mood variation in bipolar disorder," *Archives of General Psychiatry*, vol. 52, no. 11, pp. 947–959, Nov 1995. [Online]. Available: http://archpsyc.ama-assn.org/cgi/doi/10.1001/archpsyc.1995.03950230061009

[6] A. D. Krystal, "Low-dimensional chaos in bipolar disorder?" *Archives of General Psychiatry*, vol. 55, no. 3, pp. 275–276, Mar 1998. [Online]. Available: http://archpsyc.ama-assn.org/cgi/doi/10.1001/archpsyc.55.3.275

[7] S. M. Pincus, "Quantitative assessment strategies and issues for mood and other psychiatric serial study data," *Bipolar Disorders*, vol. 5, no. 4, pp. 287–294, Aug 2003. [Online]. Available: http://www.blackwell-synergy.com/links/doi/10.10340036.x

[8] ——, "Approximate entropy as a measure of system complexity," *Proc. Nat. Acad. Sci. USA*, vol. 88, pp. 2297–2301, 1991.

[9] V. K. Yeragani, R. Pohl, M. Mallavarapu, and R. Balon, "Approximate entropy of symptoms of mood: an effective technique to quantify regularity of mood," *Bipolar Disorders*, vol. 5, no. 4, pp. 279–286, Aug 2003. [Online]. Available: http://www.blackwell-synergy.com/links/doi/10.10340012.x

[10] J. M. Bopp, D. J. Miklowitz, G. M. Goodwin, W. Stevens, J. M. Rendell, and J. R. Geddes, "The longitudinal course of bipolar disorder as revealed through weekly text messaging: a feasibility study, text message mood charting," *Bipolar Disorders*, vol. 12, no. 3, pp. 327–334, May 2010. [Online]. Available: http://doi.wiley.com/10.1111/j.1399-5618.2010.00807.x

[11] M. Bonsall, S. Wallace-Hadrill, J. Geddes, G. Goodwin, and E. Holmes, "Nonlinear time series approaches in characterizing mood stability and mood instability in bipolar disorder." *Proceedings of the Royal Society B: Biological Sciences*, 2011.

[12] G. U. Yule, "On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers," *Philosophical Transactions of the Royal Society A Mathematical Physical and Engineering Sciences*, vol. 226, no. 636-646, pp. 267–298, Jan 1927. [Online]. Available: http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.1927.0007

[13] R. Brown, *Statistical forecasting for inventory control*. New York: McGraw-Hill, 1959.

[14] C. C. Holt, "Forecasting seasonals and trends by exponentially weighted moving averages," *International Journal of Forecasting*, vol. 20, no. 1, pp. 5–10, Jan 2004. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0169207003001134

[15] P. R. Winters, "Forecasting sales by exponentially weighted moving averages," *Management Science*, vol. 6, no. 3, pp. 324–342, Apr 1960. [Online]. Available: http://mansci.journal.informs.org/cgi/doi/10.1287/mnsc.6.3.324

[16] J. F. Muth, "Optimal properties of exponentially weighted forecasts," *Journal of the American Statistical Association*, vol. 55, no. 290, p. 299, Jun 1960. [Online]. Available: http://www.jstor.org/stable/2281742?origin=crossref

[17] G. Box and G. Jenkins, *Time series analysis: Forecasting and control.*, 1st ed. San Francisco: Holden-Day, 1970.

[18] R. Hyndman, A. Koehler, K. Ord, and R. Snyder, *Forecasting with Exponential Smoothing: The State Space Approach*. Springer, 2008.

[19] H. Tong and K. S. Lim, "Threshold autoregression, limit cycles and cyclical data," *Journal of the Royal Statistical Society*, vol. 42, no. 3, 1980.

[20] J. G. de Gooijer and R. J. Hyndman, "25 years of time series forecasting," *Int. J. Forecast.*, vol. 22, no. 3, pp. 443–473, 2006.

[21] J. De Gooijer, "Some recent developments in non-linear time series modelling, testing, and forecasting," *International Journal of Forecasting*, vol. 8, no. 2, pp. 135–156, Oct 1992. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/016920709290115P

[22] A. O'Hagan, "Curve fitting and optimal design for prediction." *Journal of the Royal Statistical Society*, vol. 40, no. 1, pp. 1–42, 1978.

[23] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[24] C. Rasmussen, "Evaluation of gaussian processes and other methods for non-linear regression." Ph.D. dissertation, University of Toronto, 1996.

[25] O. Stegle, S. Fallert, D. MacKay, and S. Brage, "Gaussian process robust regression for noisy heart rate data," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 9, pp. 2143–2151, Sep 2008. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4487100

[26] G. Salimi-Khorshidi, T. E. Nichols, S. M. Smith, and M. W. Woolrich, "Using gaussian-process regression for meta-analytic neuroimaging inference based on sparse observations," *IEEE Transactions on Medical Imaging*, vol. 30, no. 7, pp. 1401–1416, Jul 2011. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5723754

[27] S. Faul, G. Gregorcic, G. Boylan, W. Marnane, G. Lightbody, and S. Connolly, "Gaussian process modeling of eeg for the detection of neonatal seizures," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 12, pp. 2151–2162, Dec 2007. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4359992

[28] A. Rush, "The 16-item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression." *Biological Psychiatry*, vol. 54, no. 5, 2003.

[29] *Diagnostic and Statistical Manual of Mental Disorders: Text Revision, 4th edition.*, American Psychiatric Association, Washington, DC, 2000.

[30] E. G. Altman, D. Hedeker, J. L. Peterson, and J. M. Davis, "The altman self-rating mania scale," *Biological Psychiatry*, vol. 42, no. 10, pp. 948–955, Nov 1997. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0006322396005483

[31] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 1st ed. Cambridge University Press, 2003.

[32] C. K. Peng, S. Havlin, H. Stanley, and A. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time-series," *Chaos*, vol. 5, no. 1, pp. 82–87, 1995.

[33] M. Little, P. McSharry, I. Moroz, and S. Roberts, "Nonlinear, biophysically-informed speech pathology detection," *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006 Proceedings.: Toulouse, France*, 2006.

[34] Z. Chen, P. C. H. Ivanov, K. Hu, and H. E. Stanley, "Effect of nonstationarities on detrended fluctuation analysis." *Physical review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 65, no. 4 Pt 1, Apr. 2002.